Brazilian Journal of Biology®

ISSN 1519-6984 (Print)
ISSN 1678-4375 (Online)

Original Article

Identification of early flowering mutant gene in *Phalaenopsis amabilis* (L.) Blume for sgRNA construction in CRISPR/Cas9 genome editing system

Identificação do gene mutante de floração precoce em *Phalaenopsis amabilis* (L.) Blume para construção de sgRNA no sistema de edição de genoma CRISPR/Cas9

N. P. A. E. O. Suputri[a], I. S. Prasojo[a], L. A. T. Prabowo[a], Y. A. Purwestri[a], Purnomo[a] and E. Semiarti[a*]

[a]Universitas Gadjah Mada, Faculty of Biology, Department of Tropical Biology, Yogyakarta, Special Region of Yogyakarta, Indonesia

Abstract

Phalaenopsis amabilis (L.) Blume commonly called Moth Orchid (*Orchidaceae*) is a natural orchid species designated as the National Flower of Indonesia for its beautiful flower shape and long-lasting flowering period. Basically, *P. amabilis* has a long vegetative phase that cause late flowering, about 2 to 3 years for flowering, hence a method to shorten vegetative period is desired. The latest technological approach that can be used to accelerate flowering of *P. amabilis* is the CRISPR/Cas9 genome editing method to inactivate the *GAI* (*Gibberellic Acid Insensitive*) gene as a mutant gene that can accelerate the regulation of *FLOWERING TIME* (*FT*) genes flowering biosynthesis pathway. The approach that needs to be taken is to silence the *GAI* gene with a knockout system which begins with identifying and characterizing the *GAI* target gene in the *P. amabilis* which will be used as a single guide RNA. CRISPR/Cas9 mediated knockout efficiency is highly dependent on the properties of the sgRNA used. SgRNA consists of a target sequence, determining its specificity performance. We executed phylogenetic clustering for the *PaGAI* protein with closely related orchid species such as *Dendrobium capra, Dendrobium* cultivars and *Cymbidium sinensis*. SWISS-Model as tool webserver for protein structure homology modeling. Results show that *P. amabilis* has a specific domain with the occurrence of point mutations in the two conservative domains. Therefore, a single guide RNA reconstruction needs to be implemented.

Keywords: CRISPR/Cas9, *Gibberellic Acid Insensitive*, knockout, *Phalaenopsis amabilis*, sgRNA.

Resumo

Phalaenopsis amabilis (L.) Blume, comumente chamada de orquídea mariposa (*Orchidaceae*), é uma espécie natural de orquídea designada como a flor nacional da Indonésia por seu belo formato de flor e período de floração duradouro. Basicamente, *P. amabilis* tem uma longa fase vegetativa que causa floração tardia, cerca de 2 a 3 anos para a floração, portanto, um método para encurtar o período vegetativo é desejado. A mais recente abordagem tecnológica que pode ser utilizada para acelerar a floração de *P. amabilis* é o método de edição do genoma CRISPR/Cas9 para inativar o *GAI (Gibberellic Acid Insensitive*) que pode ser usado como um gene mutante para acelerar a regulação da floração dos genes *FLOWERING TIME (FT),* via de biossíntese. Para isto, a melhor abordagem é silenciar o *GAI* gene com um sistema *knockout* que deve ser iniciado com a identificação e caracterização do gene alvo *GAI* no *P. amabilis,* e que, posteriormente, será utilizado como um único RNA guia. A eficiência de nocaute mediada por CRISPR/Cas9 é altamente dependente das propriedades do sgRNA usados. O SgRNA consiste em uma sequência alvo, determinando seu desempenho de especificidade. Executamos agrupamento filogenético para a proteína *PaGAI* com espécies de orquídeas intimamente relacionadas, como *Dendrobium capra, Dendrobium cultivars e Cymbidium sinensis*. SWISS-Model foi utilizado como ferramenta *webserver* para modelagem de homologia de estruturas de proteínas. Os resultados mostram que *P. amabilis* possui um domínio específico com ocorrência de mutações pontuais nos dois domínios conservativos. Portanto, uma única reconstrução de RNA guia precisa ser implementada.

Palavras-chave: CRISPR/Cas9, *GAI* (*Gibberellic Acid Insensitive*), knockout, *Phalaenopsis amabilis*, sgRNA.

1. Introduction

Phalaenopsis amabilis (L.) Blume is one of orchid species, a member of the largest Angiosperm families (Family *Orchidaceae*; Genus *Phalaenopsis*) and has a high rate of speciation with a character that breeders like as a mother plant, because it has a beautiful morphology, attractive colors, and fragrant aromas (Peakall, 2007). This orchid species is a commodity traded in floriculture as cut flowers and potted flowers.

*e-mail: endsemi@ugm.ac.id
Received: September 23, 2022 – Accepted: April 12, 2023

The flowering process of orchids can be generally divided into two, *i.e.*: flowering transition and flower development (Wang et al., 2019). The flowering genes expression is regulated by key flowering genes, one of which is *PaFT (Phalaenopsis aphrodite FLOWERING LOCUS-T)* (Ziv et al., 2014).

Flowering transition is the process of organ formation in the vegetative to reproductive phase regulated by which the vegetative gene group is regulated homeobox *KNAT1*, while in contrast, the flowering gene is regulated by the *MADS-box* homeotic gene. The problem with *P. amabilis* is that the orchid requires long period to flower. The flowers bloom after a long period of about 130 weeks after sowing planting (WASP) (Semiarti et al., 2007) and can reach a diameter of 10 cm (Figure 1). The transition phase from vegetative to flowering in *Arabidopsis thaliana* starts from the activity of the *CONSTAN* gene inducing the formation of the *FT (Flowering Time locus-T)* protein with *FD (Flowering Time locus-D)* to increase the *LEAFY* and *APETALA* genes which are genes that regulate the apical stem meristem into a flower meristem (Turck et al., 2008). Research related to flowering using *PaFT* gene expression has been carried out by Semiarti et al. (2015) by overexpressing the *PaFT* gene in *P. amabilis* orchid plants which can cause the expression of the *POH1* gene (vegetative phase coding gene) to decrease so that the vegetative phase can be shortened (Semiarti et al., 2015). However, the presence of the *GAI (Gibberellic Acid Insensitive)* gene as an inhibitor can inhibit the regulation of the *PaFT* gene (Ding et al., 2013). The GAI protein is a transcription factor gene of the DELLA family as a protein repressor of the gibberellic acid pathway (Ding et al., 2013). This problem can be solved using the CRISPR/Cas9 genome editing method.

In general, many studies have been published on modified flowering times (Sun and Kao, 2018; Tong et al., 2020; Xu et al., 2020). The advantage of using CRISPR/Cas9 genome editing is the ability to edit multiple target genes simultaneously. In addition, plants with the desired characteristics can be obtained much faster than other breeding methods. However, the use of Phalaenopsis genome editing has its limitations, and the genome editing method for early flowering purposes has not been applied in the *P. amabilis* orchid species. Therefore, studies related to the regulation of early flowering by genome editing should be conducted on *P. amabilis*.

The genome editing method used by CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated 9 protein) can be found in bacteria and was derived from it. Viral memory is made possible by the CRISPR array in bacteria. The bacteria create RNA fragments from the CRISPR array that are designed to target the virus' DNA. In order to produce a crRNA complex, also known as guide RNA, CRISPR, also known as crRNA, collaborates with trans-activating CRISPR/ RNA (tracrRNA). This complex will guide the Cas protein toward a specifically marked target and will mutate the genome (El-Mounadi et al., 2020). CRISPR/Cas9 has three main components consisting of sgRNA as a guide, PAM as an indicator of the target cutting position and Cas9 as a cutting shear. The Cas9 and sgRNA complexes will cut the target gene, followed by a repair system with two alternative genome repair mechanisms, *i.e.*: Non-Homologous End-Joining (NHEJ) or Homology Directed Repair (HDR). The latest technological approach that *P. amabilis* can use is the CRISPR/Cas9 genome editing method to activate the *GAI* gene.

The approach that needs to be taken is to silence the *GAI* gene with a knockout system which begins with identifying and characterizing the *GAI* target gene in the genus *Phalaenopsis* which will be used as a single guide RNA. The CRISPR/Cas9 mediated knockout efficiency is highly dependent on the properties of the sgRNA used. Single guide RNA consists of a sequence of targeting, determining its specificity, followed by RNA at the 3′ ends (Wang et al., 2016). Wang et al. (2016) show that the presence of different motives in sgRNA results in 10 times lower to reduce in the knockout frequency of the gene. Mechanistically, the cause of low efficiency depends largely on the motif used and must be specified on the target genome used. This sequence motif is relevant for future sgRNA design approaches and interaction studies of CRISPR/Cas9 performance effectiveness influences.

Identification is carried out based on the phenomenon of flowering time between different genera of the *Orchidaceae* family. This allows for genotypical character differences in the process of flowering meristem formation which is closely related to the regulation of the key genes of *DELLA* inhibitors in the regional structure of the N and C terminals. This study characterizes target genes based on 1) identification of *GAI* gene transcription and protein domain analysis of *GAI* genes, 2) characterization of *GAI* gene physio-chemical, 3) phylogenetic analysis and prediction of orthologs genes, 4) GAI protein homolog modeling was carried out based on the analysis of secondary protein structure (2D) and tertiary protein structure (3D) which showed the presence of alpha helix and beta-turns in the protein structure. The aim of this study was to determine the protein structure of the *GAI* gene in *P. amabilis* for the basic steps of making sgRNA in genome editing using the CRISPR/Cas9 method.

Figure 1. Habitus growth and development of *P. amabilis*. Seedlings on 3–36 WASP. adult plants on 130 WASP and a flower. Scale bars with 5 mm for 3–18 WASP. 1 cm for 36 WASP. 5 cm for 130 WASP and 2 cm for flower.

2. Material and Methods

2.1. DNA extraction and gene amplification using GAI (Gibberellic Acid Insensitive) degenerated primers of Phalaenopsis amabilis (L.) Blume

Genomic DNA isolation of *the P. amabilis* was conducted using a protocol from the Murray and Thompson method (Murray and Thompson, 1980). Three months old plantlet being sampled. The isolated gDNA of *P. amabilis* was amplified with *DegGAI* primers to get the target gene. The primer sequences utilized are shown in Table 1 and Figure 2.

Amplification of the *DELLA* (*GAI*) gene was performed using a pair of *DegGAI* degenerated primer. Amplification was performed using Bioline MyTaq™ (UK) Mix PCR. The mixture for PCR reaction is shown in Table 2.

The component is homogenized using micropipettes, then incubated using a thermal cycler machine at a temperature of 98°C, 10 seconds; 60°C, 15 seconds; 72°C, 50 seconds for 40 cycles. The amplification results were electrophoresis and visualized using a UV trans-illuminator.

2.2. Identification of GAI genes transcription and analysis protein domain of GAI genes

To identify *GAI* gene transcription, GAI protein sequences from *Dendrobium capra* were used as query sequences NCBI database by BLASTn or Basic Local Alignment Search Tool based on nucleotide performed on *Phalaenopsis equestris* transcriptomes in Orchidstra 5.0 (Chao et al., 2017). MultAlin tool (Corpet, 1988) was employed to identify the DNA-binding N-terminus and C-terminus regions. The online server MEME suite (Bailey et al., 2009) was used to identify the conserved motifs of GAI protein.

2.3. Physio-chemical characterization of GAI proteins structural

The Expasy-ProtParam server (Gasteiger et al., 2007) was utilized to decide the assumed physio-chemical properties in parameters physio-chemical properties in parameters of molecular weight, aliphatic index, instability index, isoelectric point, and large average hydrophobicity (GRAVY)). The localization of subcellular proteins was predicted using CELLO v.2.5 (Yu et al., 2006) and PSORT (Horton et al., 2007). For the presence of the signal peptide sequence and the helical transmembrane region, the signal P. 4.0 (Petersen et al., 2011) and TMHMM v. 2.0 (Petersen et al., 2011) was used to analyze protein sequences.

2.4. Phylogenetic analysis of GAI proteins

Full-length protein arrangements using the integrated MUSCLE program and further analyzed using the MEGAXI tool (Kumar et al., 2018) to set up phylogenetic connections between all GAI proteins. Phylogenetics was built concurring to the greatest probability utilizing Jones-Taylor-Thornton (JTT) with 5-fold gamma dissemination and a demonstration with a 1000-fold gamma distribution.

2.5. Molecular modelling of GAI proteins

Selected GAI protein sequences were analyzed for prediction using the SOPMA secondary structure prediction tool (Sapay et al., 2006) to predict secondary structures, alpha helices, random coils, and beta turns. SWISS-MODEL is a protein structural homology model that can simulate 3D structures based on references in the Protein Data Bank (PDB) database (Studer et al., 2020).

Figure 2. Illustrated map of degenerated primer of *GAI* gene. the gene loci based on the sequences from gene bank.

Table 1. Degenerate primer of *GAI* gene sequences.

	Position	Sequence (5'-3')	Amplicon
Deg*GAI*	Forward	ATGAAGACAGAGCACCTTGAGAGC	574 bp
	Reverse	CTACAGGATTAGTAGATGATACCTCGGG	

Table 2. PCR Components for *GAI* gene amplification in *P. amabilis*.

No.	Material	Composition
1.	MyTaq™ Mix PCR Bioline	12.5 µl
2.	DegGAI Primer forward (10 µM)	1 µl
3.	DegGAI Primer reverse (10 µM)	1 µl
4.	gDNA Templates (50 ng)	1 µl
5.	ddH_2O	up to 25 µl

QMEAN (Qualitative Model Energy Analysis) and QMQE (Global Model Quality Estimate) values were evaluated to measure the global scale by Z-Score as well as model evaluation by local quality estimation is also used to quantify protein quality. GAI protein sequences in 3D were analyzed any computational tools such as CHIMERA X (Goddard et al., 2018), AutoDock 4.2 (Morris et al., 2009), PyRx docking tools (Dallakyan and Olson, 2015), PyMol visualization tools (Schrödinger, 2020).

3. Results

3.1. Analysis of GAI gene in P. amabilis using GAI Deg primers

The amplified DNA fragment of *GAI* gene with *DegGAI* primers resulted in a single band with 576 bp length, as seen in Figure 3. The nucleotide sequence was translated into protein sequence thus analyzed *in silico* to determine the character of the GAI protein structure in *P. amabilis*.

3.2. Identification of GAI genes transcription and analysis protein domain of GAI genes

The PaGAI amino acid sequence analysis shows 2 regions of conservative amino acid motifs. There are several gaps present in this alignment. Variations of motifs in the amino acid level have the potential for structural and functional gene changes. Amino acid level polymorphism was analyzed with DNASP. Results of the analysis showed that the orchids *C. sinensis, D. hybrid* and *D. capra* had a DEQLA motif while *P. amabilis* had a DELLA motif. The polymorphisms also occur in GRAS domain PFYRE motifs (Figures 4 and 5).

Polymorphism analysis showed that in the DELLA domain, the *P. amabilis* orchid one of the amino acids Glutamine (Q) changes to the amino acid Leucine (L). Hence, the initial motif of DEQLA changed to DELLA (Figure 5A). Motif changes also occur in *P. amabilis* in the GRAS domain in amino acids from Glutamic acid (E) to the Glutamine (Q), Serine (S) to Arginine (A), Glutamic acid (E) to the amino acid Glutamine (Q) Cysteine (C) to Histidine (H) (Figure 5B).

Result of the analysis of the motif (Figure 6) shows that there were six motifs on the alignment. The data also show that *P. amabilis* has no consensus in motif number 3, which also shows that the *GAI* gene in *P. amabilis* has a specific domain compared to the other three orchids. This is supportive for subsequent research in terms of constructing sequences specifically using specific primers in designing sgRNA. According to Doench et al. (2016) location and sequence are important considerations for designing gRNAs. The creation of gRNA sequences with high activity and reduced off targets is crucial. These considerations are important for CRISPRa and CRISPRi. The sgRNA sequence must be determined using the target sequence with the right coding sequence through the stages of identification analysis and characterization of the target to be functionally validated, to provide further utility to the database, computationally designed sequences are also obtained from various sources, mainly focused on protein-coding genes (Doench et al., 2016; Chari et al., 2015).

3.3. Phylogenetic analysis

The phylogenetic analysis results in Figure 7 shows that *the GAI* gene has a difference index of 28% in *P. amabilis*. This indicates that *P. amabilis* has specific characteristic motif in the *GAI* genes.

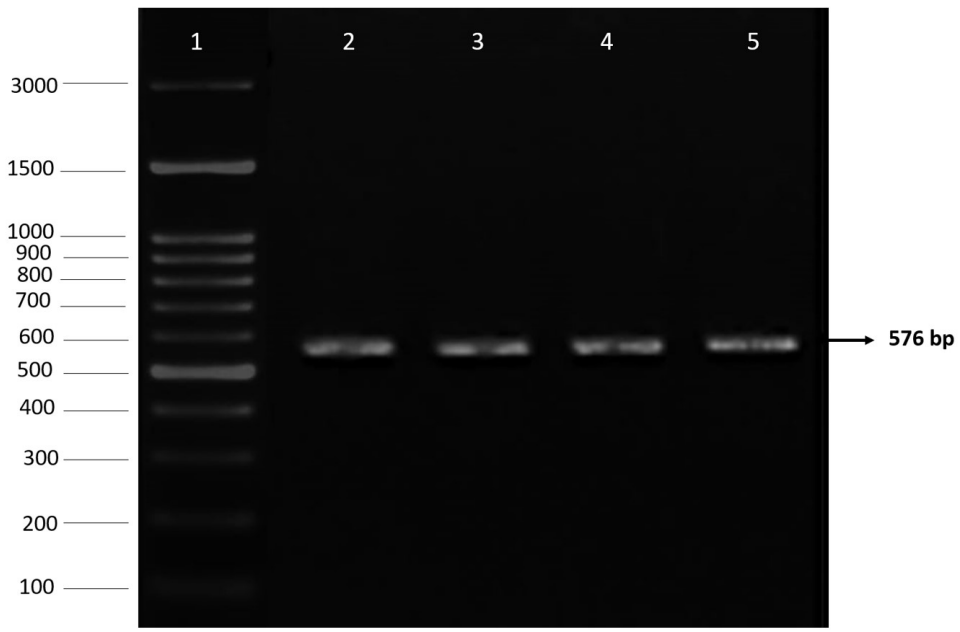

Figure 3. Amplified fragment of *GAI genes* in *P. amabilis* using *DegGAI* Primers. Lane 1: 100 bp DNA marker; Lane 2-5: amplicon of *GAI* gene fragments 1-4.

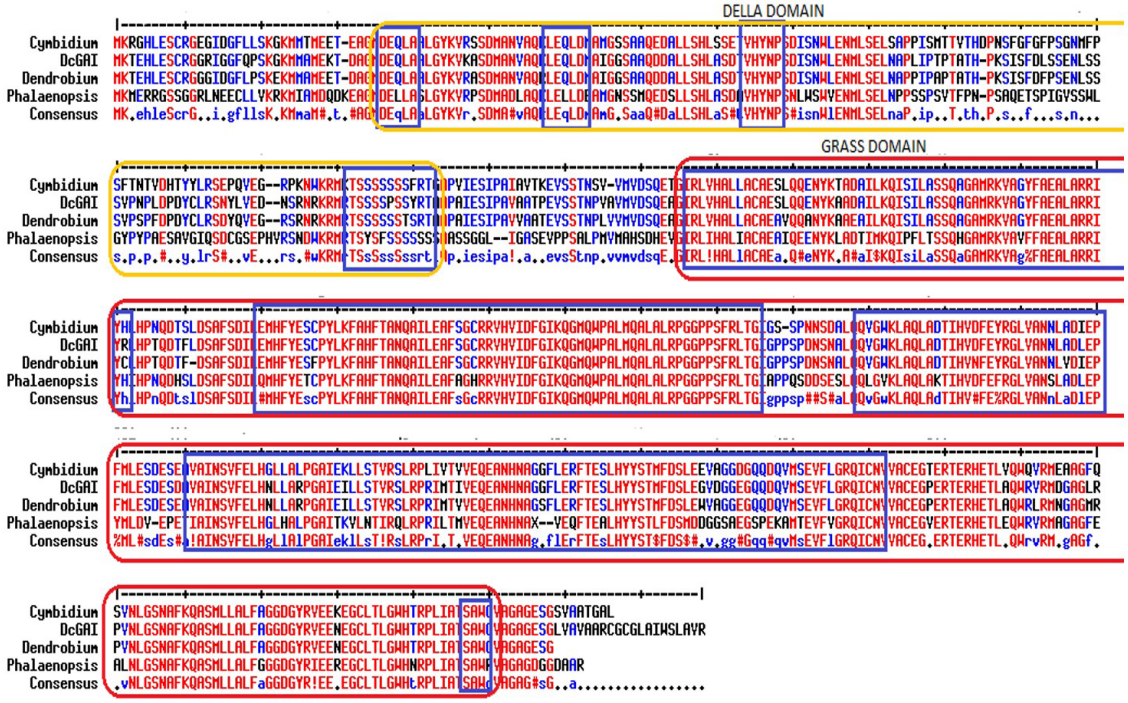

Figure 4. Partial sequence alignment of GAI protein and other orchids. The constituent amino acid sequences from four orchid plants of the *Orchidaceae* family which includes *GAI* genes and protein constituent amino acids for the DELLA and GRAS domains indicate several gaps formed in the alignment with the presence of some amino acid that are not present in other orchids. Differences in amino acid arrangements affect protein function.

A. DELLA conservative domain

```
                ....|
Cymbidium       DEQLA
DcGAI           DEQLA
Dendrobium      DEQLA
Phalaenopsis    DELLA
```

B. GRAS conservative domain

```
      ....|....| ....|....| ....|....| ....|....| ....|....|
                          10         20         30         40
     50
Cymbidium       EALAREMHFY ESCPYLKFAH FTANQAILEA FSGCRRVHVI
DFGIKQGMQW
DcGAI           EALAREMHFY ESCPYLKFAH FTANQAILEA FSGCRRVHVI
DFGIKQGMQW
Dendrobium      EALAREMHFY ESCPYLKFAH FTANQAILEA FSGCRRVHVI
DFGIKQGMQW
Phalaenopsis    EALARQMHFY ETCPYLKFAH FTANQAILEA FAGHRRVHVI
DFGIKQGMQW


                ....|....| ....|....|.
                          60         70
Cymbidium       PALMQALALR PGGPPSFRLT G
DcGAI           PALMQALALR PGGPPSFRLT G
Dendrobium      PALMQALALR PGGPPSFRLT G
Phalaenopsis    PALMQALALR PGGPPSFRLT G
```

Figure 5. Analysis of amino acid polymorphism's structure DELLA domain and GRAS domain.

3.4. Physio-chemical characterization of GAI proteins

The physio-chemical properties of GAI proteins are estimated and listed (Table 3). The average length of the GAI protein peptide in *C. sinensis* is 581 aa, in *Dendrobium* cultivar is 574 aa, in *D. capra* is 577 aa and in *P. amabilis* is 576 aa. The molecular weight of the proteins of the four orchids has values that are in the range of 62-63 kDa. The isoelectric point with a range of 5.11 – 5.35 and the average aliphatic index is 81-83. The Instability Index is between 50,57. All GAI proteins have negative GRAVY (grand average value of hydropathy), characteristic of nucleotide-binding proteins. All GAI proteins are thought to be localized in the nucleus (Table 3).

3.5. Homology modelling of proteins

SWISS-MODEL homology modelling program appeared that four proteins had a comparable structure using GMQE and QMEAN scores. All the structures had been chosen by a few markers such as high determination (<2.2Å) and scope level. At that point the comes about of the investigation were sorted agreeing to the quality level of the structure (Table 4).

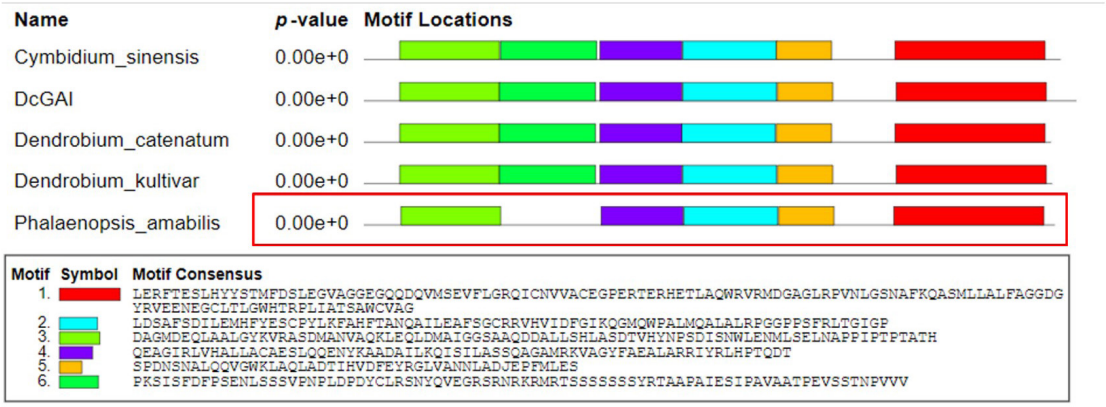

Figure 6. Protein motif analysis of GAI using MEME website tools.

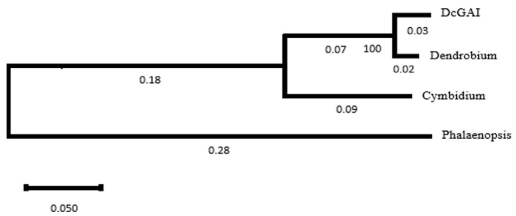

Figure 7. Phylogenetic analysis of nucleotide sequences of *the GAI* gene using MEGA 11 software using the Jones Taylor Thornton model on the gamma distribution.

Table 3. Physio-chemical characterization of GAI proteins.

Protein	Peptide Length	Molecular Weight	Isoelectric Point	Instability Index	Aliphatic Index	GRAVY	Localization	Signal Peptide	Trans-Membrane Domain
Cymbidium	581	63.4	5.11	48.44	81.82	-0.137	nucleus	0	0
D. cultivar	574	62.5	5.15	50.74	83.89	-0.205	nucleus	0	0
D. capra	577	62.3	5.18	48.99	83.55	-0.194	nucleus	0	0
P. amabilis	576	63.1	5.35	54.12	81.89	-0.215	nucleus	0	0

Table 4. Homology modelling of GAI proteins were described by SWISS model program.

Protein	GMQE Value	QMEAN Value	Sequences Identity	Sequences Similarity	Coverage	Sources	Method by	Resolution
Cymbidium sinensis	0.44	-0.66	51.76	0.43	0.13	HHblits	X-ray	1.8Å
Dendrobium cultivar	0.11	-0.62	51.16	0.47	0.15	BLAST	X-ray	1.8Å
Dendrobium capra	0.09	-0.39	60.49	0.47	0.14	BLAST	X-ray	1.8Å
Phalaenopsis amabilis	0.1	-0.67	53.57	0.44	0.14	BLAST	X-ray	1.8 Å

QMEAN (Qualitative Model Energy Analysis) and GMQE (Global Model Quality Estimate).

All GAI amino acid sequences have the same protein template model. Specific binding affinities (kcal/mol) are used to determine the optimal docking position among the five possible docking positions, and the one with the lowest free energy has the lowest binding affinity to the active site considered the best conformation of the ligand of modeled protein. In this case, the targeting ligand GA3 with the complex GAI-GID1 was combined with the ligands in Table 5 and modeled targets, software such as CHIMERA X (Goddard et al., 2018), AutoDock 4.2 (Morris et al., 2009), PyRx docking tools (Dallakyan and Olson, 2015), PyMol visualization tools (Schrödinger, 2020) were used in this research.

Four GAI amino acid sequences were analyzed for structural simulation and homological modelling and comparative graph using non redundant set of PDB structures and local quality estimates. (Figures 8 and Table 5). The secondary structure consists of random coils and an alpha-helices structure analysis (Figure 9). Simulation of the three-dimensional structures of GAI-GID1 protein interaction with GA3 ligands (Figure 10A). The target interacting residues at its active site exhibited four distinct types of interaction encompassing hydrogen bonding [Tyr134], Hydrophobic interactions [Phe27], [Ile24], [Lys28], Van der Waals interactions [Arg35], [Tyr245], [Ser123], [Arg251], [Gly122], [Ser198], [Leu330], [Ser127], [Ile133], [Gly327], [Val326], [Phe245], [Asp250], [Thr247], [Val246], and Carbon Hydrogen Bond [His126] shown in Figure 10B. Obtained Root Mean Square Deviation (RMSD) on value and affinity binding. RMSD describes the distance value of an atom at one conformation with the nearest atom having the same type as that atom in another conformation. In contrast, the RMSD values of the ligand are rather high (ranging to a maximum of 9 Angstrom) (Sargsyan et al., 2017). A total of five position of docking simulations were obtained as results of docking studies of modelled target and the ligand GA3 using AutoDock Vina 1.5.6.). The RMSD values of the five target proteins indicate that the RMSD values are small, therefore indicates that the estimated position of the ligand is closer with the natural mechanism of *GAI* gene

3.6. Designing a single guide RNA

This research began with designing a single guide RNA sequence isolated from *P. amabilis*. Furthermore, the sgRNA is cut and ligated into the pRGEB32 plasmid and inserted into *Escheria coli* strain DH5α for cloning. The stability of sgRNA in the pRGEB32 plasmid was analyzed using Polymerase Chain Reaction (PCR). The gRNA cloning site is located in upstream of U3 promoter and BsaI restriction enzyme. The truncation point used to insert the SgRNA of the BsaI target. Figure 11 shows the backbone (A) and (B) are the result of the sgRNA design. The plasmid construct has endogenous Cas9 with rice snoRNA U3 promoter to express the gRNA. Target-specific spacers are designed and cloned into the vector background by digesting the plasmid with the *Bsa*I enzyme.

4. Discussion

Plant growth and development are regulated by various environmental factors and hormones during the process of growth and development. Gibberellins (GA) are important regulators in seed germination, leaf development, root and stem growth, floral organ development, and fruit maturation (Phokas and Coates, 2021). DELLA genes included the GAI, RGA, RGL1, RGL2 and RGL3 genes (Bolle et al., 2000).

DELLA proteins are located in the plant nucleus and their conserved terminal C-GRAS domains are predominantly composed of two leucine heptapeptide repeats (LHRI and LHRII) and three conserved repeat motifs (VHIID, PFYRE and SAW). GA stimulates the formation of GA-GID1-DELLA complexes, which are further targeted for degradation at the 26S proteasome (Cheng et al., 2021). The amino acid sequence of DELLA protein is also divided into different domains. The N-terminus is a highly conserved DELLA sequence, and the flanking domain is a highly conserved TVHYNP region that helps DELLA and GID1 proteins interact with each other.

The DELLA domain of the orchid *P. amabilis*, one of the amino acids Glutamine (Q) transforms to the amino acid Leucine, according to the results of the polymorphism analysis, amino acid alignment. Therefore, the original pattern of DEQLA became DELLA (Figure 5A). In *P. amabilis*, motif alterations also occur in the GRAS domain, where the amino acid from Glutamic acid (E) to the Glutamine (Q), Serine (S) is changed to Arginine (A), and mutations also occur in the amino acid Cysteine (C) is changed to Histidine (H) (Figure 5B). According to the findings of the motif analysis, six motifs on the four orchids were chosen (Figure 6). The consensus sequences for each domain are used to derive the motif analysis.

Table 5. Binding affinity of different dock poses.

Mode	Affinity	Distance from best mode	
	(kcal/mol)	R.M.S.D	R.M.S.D
		lower bound	upper bound
1	-11	0	0
2	-10.2	1.657	3.667
3	-10.2	2.031	4.522
4	-10	1.525	5.254
5	-9.6	1.69	2.027

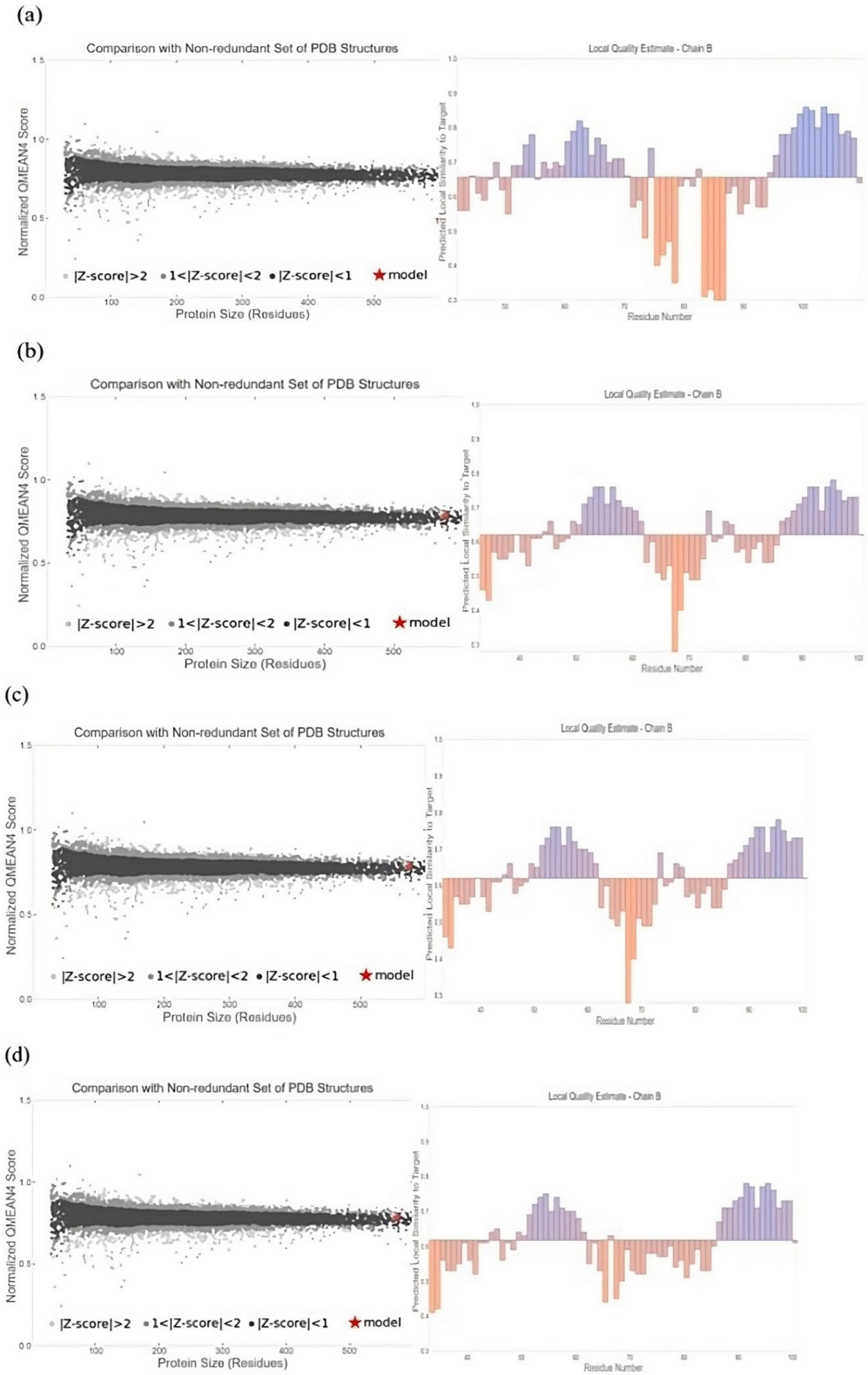

Figure 8. Comparative graph using non-redundant set of PDB structures and local quality estimates in of (a) *C. sinensis*. (b) *D. cultivar*. (c) *D. capra*. and (d) *P. amabilis* proteins. *C. sinensis* (Figure 8) had the highest GMQE score and medium-quality QMEAN Z score. All GAI proteins were of low-quality structures globally according to local quality estimates with Z-score value <1.

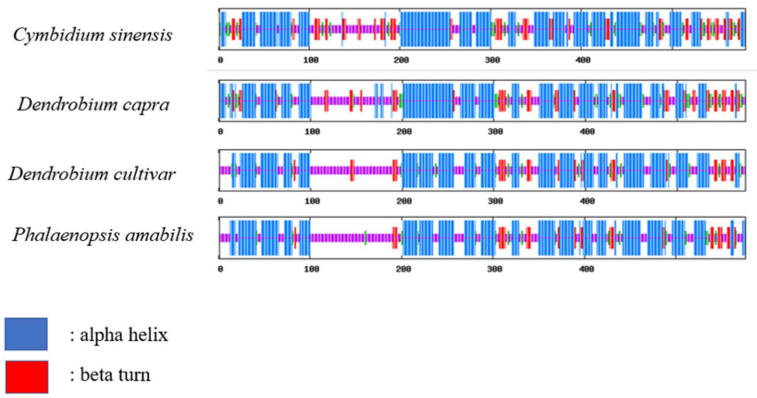

Figure 9. GAI protein structure analysis. Two-dimensional structure of *C. sinensis*. *Dendrobium* cultivar. *D. capra* and *P. amabilis A.*

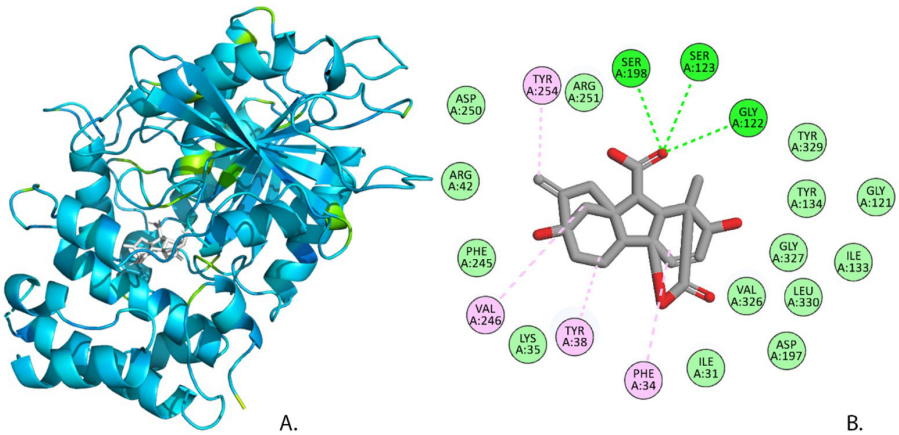

Figure 10. (A) GAI - GID1 docked in the binding pocket of GA3. (B) 2D representation of the interaction between GAI-GID1 and GA3.

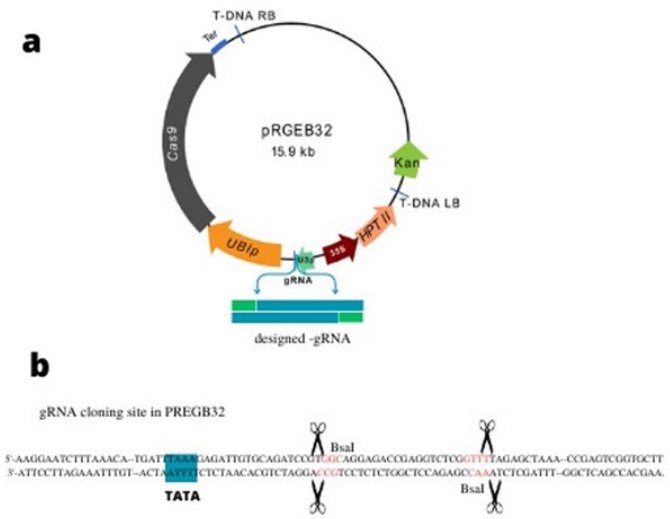

Figure 11. Development of pREGB32 carrying gRNA arrangement. a. Schematic figure of the pRGEB32 vector carrying CRISPR/Cas9 with designed-gRNA cloning location. b. The entry location of sgRNA fragments is in the upstream U3 section. where in the cutting process using the enzyme *BsaI*. The plasmid originates from the collaboration between Universitas Gadjah Mada and Nagoya University (2017-2020) through the partnership of the Japan Society for the Promotion of Science Project.

The statistics also indicate that *P. amabilis* lacks a consensus motif for motif No. 3. This indicates that, in contrast to the previous three orchids, *P. amabilis GAI* gene has a particular domain. This is supported by phylogenetic tree analysis data where evolutionarily *P. amabilis* has characteristic traits outside the group that allow for a lot of genetic variation which affects the regulatory system of key genes, especially the DELLA family genes. The results of the phylogenetic analysis in Figure 7 show that the *P. amabilis* GAI gene has a dispersion index of 28%. The amino acid spacing between TVHYNP and DELLA domains is also important for GA signal reception. Our data suggest that polymorphisms in this region and motifs occur due to evolutionary factors between different genera *Phalaenopsis, Cymbidium and Dendrobium* and different environmental adaptability and thus may influence morphogenesis in general. According to Vaidya et al. (2018), genera and environmental factors can change one or more amino acids. Biotic and abiotic environmental conditions change continuously across biological systems that contribute to changes in biological mechanisms including polymorphisms.

The physio-chemical properties of GAI proteins have been calculated in molecular weight, aliphatic index, instability indexes, and GRAVY. This demonstrates that GAI has a physically similar character on all four orchids. This is shown by physio-chemical properties where the GAI protein in the four orchid species has an average peptide length of 574-581 aa, molecular weight ranged from 62-63 kDa, isoelectric point in the range of 5.11 – 5.35 and an aliphatic index of 57.58, the index of instability ranges between 81.82 and 83.89, GRAVY is negative. Homology demonstrating plays a fundamental part in protein structure, which is essential for understanding the component of protein function. The secondary structure consists of random coils and alpha helix proteins regions which are seen in Figure 8. The SWISS-MODEL modelling program showed that the four proteins shared a similar structure with the X-ray method. All structures were selected based on several criteria such as high resolution (<2.2 Å) and close coverage value. Comparison non-redundant of data collection of PDB structures of four orchids (Figure 8) that is not excessive shows that the data obtained from and all of GAI proteins were of low-quality structures globally according to local quality estimates with Z-score value <1. This shows the PDB templates available on all four proteins show similar modeling characters. And also the 2D protein data also shows a high similarity in structure, where the GAI protein is predominantly characterized by random coil and alpha helix structures are seen in Figure 9.

The three-dimensional analysis method with a docking approach is an activity to bring the prediction of results closer and the similarity of ligand-protein interactions with native. Binding affinity is a measure of the ability protein binding with a receptor. The results of docking between ligands and receptors are obtained from the conformation of ligands with the smallest energy. The best binding affinity value with the smallest binding affinity value is shown in model one with a binding affinity value of -11 kcal/mol (Table 4). According to Liu et al. (2020), the lower bond energy value indicates the stronger affinity of the compound to the target protein while the higher energy value indicates the lower affinity of the compound. The value of the bond energy also depends on the native ligand to be compared. Affinity is the ability to bind to a target protein. The bond affinity strength of target protein ligands is influenced by free bond energy, surface interactions, and intermolecular interactions. The lower the free bond energy value, the higher the ligand and receptor bond affinity value. The low free bond energy value is able to bind to the target protein strongly and raises the potential for biological activity (Liu et al., 2020). The interaction of ligands with enzymes can also interact through hydrogen bonds and Van Der Walls bonds. This is by the interaction data of the GAI-GID1 complex protein which have four protein interaction bonds, consisting of Hydrogen bonding interactions [Tyr134], Hydrophobic interactions [Phe27], [Ile24], [Lys28], Van der Waals interactions [Arg35], [Tyr245], [Ser123], [Arg251], [Gly122], [Ser198], [Leu330], [Ser127], [Ile133], [Gly327], [Val326], [Phe245], [Asp250], [Thr247], [Val246], and Carbon Hydrogen Bond [His126] shown in Figure 10B.

According to Chen et al. (2016), there are four types of interactions namely hydrogen bonds, Van Der Walls interactions, electrostatic forces, and hydrophobic interactions are the main binding forces that affect the interaction of various substances with macromolecules. The more interactions between the GAI protein and amino acid residues the better the interaction prediction. Nugraha et al. (2018) state that the lower the energy of the ligand bond with the target protein, the stronger the bond formed between the ligand and the target protein and likewise the affinity of the ligand for the target protein which is directly proportional to the strength and stability of the bond. All five GAI proteins from 5 orchids showed small RMSD. The data indicates that the estimated position of the ligand is closer to the mechanism of the *GAI* gene and the key stage for GA sensitivity in the flowering period in the conformation of the origin in plants (Fukazawa et al., 2017).

GAI proteins were found in these four orchids, even though physio-chemical analysis and homologous modelling of secondary and triplet structures revealed unimpressive results. Based on differences in mutation variations and one conservative motif that is lost based on an analysis of the structure of the motif in amino acids, the building of sgRNA reconstruction is carried out. The design of the sgRNA is extremely important for the CRISPR-based screens. To direct the Cas9 proteins to genomic targets in a CRISPR system, complex sgRNA and PAM act as to the targeted DNA sequence. Based on the alignment of spacer sequences to the genome, several researchers have examined the impacts of sgRNA that are off target (Tsai et al., 2015). However, DNA sequence also affects how well sgRNA works at the on-target genomic loci. It is well known that the Cas9-DNA binding and cut in a CRISPR/Cas9 system require the Protospacer Adjacent Motif (PAM) carrying NGG consensus (Sternberg et al., 2014).

5. Conclusions

Our analysis of protein domains, motifs, homological modelling, and phylogenetic relationships illustrates that GAI are genes with a conservative protein structure at the sequence and structural levels. The GAI protein of *P. amabilis* has a specific amino acid sequence and consists of only 5 protein motifs.

The GAI protein of *P. amabilis* has a specific amino acid sequence and consists of only 5 protein motifs. Based on phylogenetic trees, *P. amabilis* is in an outgroup clade which is evolutionarily a species that diverged from all in group taxa before it diverged from each other and has specific character. Based on two-dimensional protein structure analysis, *P. amabilis* GAI has the number of random coils and alpha helix proteins regions in its protein sequence. Based on three-dimensional protein structures analysis, GAI-GID1 complex with GA3 ligands with a good conformation level of -11 kcal/mol binding affinity. The specific sequence of *GAI P. amabilis* is designed as sgRNA and inserted into the plasmid pRGEB32.

Acknowledgements

This work was supported by The Ministry of Education, Culture, Research and Technology for Fellowship Research Grant of Research PMDSU (Contract No: 089/E5/PG.02.00. PT/2022 1982/UN1/DITLIT/Dit-Lit/PT.01.03/2022) to Endang Semiarti as the Principal Investigator and Ni Putu Ayu Erninda Oktaviani Suputri as Doctoral Student of Faculty of Biology UGM, Yogyakarta, Indonesia.

References

BAILEY, T.L., BODEN, M., BUSKE, F.A., FRITH, M., GRANT, C.E., CLEMENTI, L., REN, J., LI, W.W. and NOBLE, W.S., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, vol. 37, no. Web Server, pp. W202-W208. http://dx.doi.org/10.1093/nar/gkp335. PMid:19458158.

BOLLE, C., KONCZ, C. and CHUA, H., 2000. PAT1, a new member of the GRAS family, is involved in phytochrome A signal transduction. *Genes & Development*, vol. 14, no. 10, pp. 1269-1278. http://dx.doi.org/10.1101/gad.14.10.1269. PMid:10817761.

CHAO, Y.-T., YEN, S.-H., YEH, J.-H., CHEN, W.-C. and SHIH, M.-C., 2017. Orchidstra 2.0. A transcriptomics resource for the orchid family. *Plant & Cell Physiology*, vol. 58, no. 1, p. e9. http://dx.doi.org/10.1093/pcp/pcw220. PMid:28111366.

CHARI, R., MALI, P., MOOSBURNER, M. and CHURCH, G., 2015. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nature Methods*, vol. 12, no. 9, pp. 823-826. http://dx.doi.org/10.1038/nmeth.3473. PMid:26167643.

CHEN, D., OEZGUEN, N., URVIL, P., FERGUSON, C., DANN, S.M. and SAVIDGE, T.C., 2016. Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Science Advances*, vol. 2, no. 3, p. e1501240. http://dx.doi.org/10.1126/sciadv.1501240. PMid:27051863.

CHENG, J., MA, J., ZHENG, X., LV, H., ZHANG, M., TAN, B., YE, X., WANG, W., ZHANG, L., LI, Z., LI, J. and FENG, J., 2021. Functional analysis of the *Gibberellin* 2-oxidase gene family in peach. *Frontiers in Plant Science*, vol. 12, p. 619158. http://dx.doi.org/10.3389/fpls.2021.619158. PMid:33679834.

CORPET, F., 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Research*, vol. 16, no. 22, pp. 10881-10890. http://dx.doi.org/10.1093/nar/16.22.10881. PMid:2849754.

DALLAKYAN, S. and OLSON, A.J., 2015. Small-molecule library screening by docking with PyRx. *Methods in Molecular Biology*, vol. 1263, pp. 243-250. http://dx.doi.org/10.1007/978-1-4939-2269-7_19. PMid:25618350.

DING, L., WANG, Y. and YU, H., 2013. Overexpression of DOSOC1, an ortholog of arabidopsis SOC1, promotes flowering in the orchid dendrobium Chao praya smile. *Plant & Cell Physiology*, vol. 54, no. 4, pp. 595-608. http://dx.doi.org/10.1093/pcp/pct026. PMid:23396600.

DOENCH, J.G., FUSI, N., SULLENDER, M., HEGDE, M., VAIMBERG, E.W., DONOVAN, K.F., SMITH, I., TOTHOVA, Z., WILEN, C., ORCHARD, R., VIRGIN, H.W., LISTGARTEN, J. and ROOT, D.E., 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, vol. 34, no. 2, pp. 184-191. http://dx.doi.org/10.1038/nbt.3437. PMid:26780180.

EL-MOUNADI, K., MORALES-FLORIANO, M.L. and GARCIA-RUIZ, H., 2020. Principles, applications, and biosafety of plant genome editing using CRISPR-Cas9. *Frontiers in Plant Science*, vol. 11, p. 56. http://dx.doi.org/10.3389/fpls.2020.00056. PMid:32117392.

FUKAZAWA, J., MORI, M., WATANABE, S., MIYAMOTO, C., ITO, T. and TAKAHASHI, Y., 2017. DELLA-GAF1 complex is a main component in gibberellin feedback regulation of GA20 oxidase 2. *Plant Physiology*, vol. 175, no. 3, pp. 1395-1406. http://dx.doi.org/10.1104/pp.17.00282. PMid:28916594.

GASTEIGER, E., HOOGLAND, C., GATTIKER, A., DUVAUD, S., WILKINS, M. and APPEL, R., 2007. Protein identification and analysis tools on the ExPASy server. In: J.M. WALKER, ed. *The proteomics protocols handbook*. Totowa: Humana Press, pp. 571-607. http://dx.doi.org/10.1385/1-59259-890-0:571.

GODDARD, T.D., HUANG, C.C., MENG, E.C., PETTERSEN, E.F., COUCH, G.S., MORRIS, J.H. and FERRIN, T.E., 2018. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Science*, vol. 27, no. 1, pp. 14-25. http://dx.doi.org/10.1002/pro.3235. PMid:28710774.

HORTON, P., PARK, K.-J., OBAYASHI, T., FUJITA, N., HARADA, H., ADAMS-COLLIER, C.J. and NAKAI, K., 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Research*, vol. 35, no. Web Server, pp. W585-W587. http://dx.doi.org/10.1093/nar/gkm259. PMid:17517783.

KUMAR, S., STECHER, G., LI, M., KNYAZ, C. and TAMURA, K., 2018. MEGA XI: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, vol. 35, no. 6, pp. 1547-1549. http://dx.doi.org/10.1093/molbev/msy096. PMid:29722887.

LIU, Y., GRIMM, M., DAI, W.T., HOU, M.C., XIAO, Z.X. and CAO, Y., 2020. CBDock: a web server for cavity detection-guided protein–ligand blind docking. *Acta Pharmacologica Sinica*, vol. 41, no. 1, pp. 138-144. http://dx.doi.org/10.1038/s41401-019-0228-6. PMid:31263275.

MORRIS, G.M., HUEY, R., LINDSTROM, W., SANNER, M.F., BELEW, R.K., GOODSELL, D.S. and OLSON, A.J., 2009. Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. *Computers & Chemistry*, vol. 30, no. 16, pp. 2785-2791. http://dx.doi.org/10.1002/jcc.21256.

MURRAY, M.G. and THOMPSON, W.F., 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, vol. 8, no. 19, pp. 4321-4326. http://dx.doi.org/10.1093/nar/8.19.4321. PMid:7433111.

NUGRAHA, W., SUWARTAWAN, W., PRAYOGA, A., LAKSMIANI, L., PUTRA, P. and ANI, S., 2018. Potensi Brazilein Potensi Brazilein dari Kayu Secang (Caesalpinia Sappan L.) sebagai Agen Depigmentasi Kulit secara In Silico. *Jurnal Farmasi Udayana*, vol. 7, no. 1, pp. 1-6. http://dx.doi.org/10.24843/JFU.2018.v07.i01.p02.

PEAKALL, R., 2007. Speciation in the Orchidaceae: confronting the challenges. *Molecular Ecology*, vol. 16, no. 14, pp. 2834-2837. http://dx.doi.org/10.1111/j.1365-294X.2007.03311.x.

PETERSEN, T.N., BRUNAK, S., VON HEIJNE, G. and NIELSEN, H., 2011. Signal, P 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, vol. 8, no. 10, pp. 785-786. http://dx.doi.org/10.1038/nmeth.1701. PMid:21959131.

PHOKAS, A. and COATES, J., 2021. Evolution of DELLA function and signaling in land plants. *Evolution & Development*, vol. 23, no. 3, pp. 137-154. http://dx.doi.org/10.1111/ede.12365. PMid:33428269.

SAPAY, N., GUERMEUR, Y. and DELÉAGE, G., 2006. Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. *BMC Bioinformatics*, vol. 7, no. 1, p. 255. http://dx.doi.org/10.1186/1471-2105-7-255. PMid:16704727.

SARGSYAN, K., GRAUFFEL, C. and LIM, C., 2017. How molecular size impacts RMSD applications in molecular dynamics simulations. *Journal of Chemical Theory and Computation*, vol. 13, no. 4, pp. 1518-1524. http://dx.doi.org/10.1021/acs.jctc.7b00028. PMid:28267328.

SCHRÖDINGER, 2020 [acessed 12 April 2023]. *PyMOL* [software]. New York: Schrödinger Available from: http://www.pymol.org/pymol

SEMIARTI, E., INDRIANTO, A., PURWANTORO, A., ISMININGSIH, S., SUSENO, N., ISHIKAWA, T., YOSHIOKA, Y., MACHIDA, Y. and MACHIDA, C., 2007. Agrobacterium-mediated transformation of the wild orchid species Phalaenopsis amabilis. *Plant Biotechnology*, vol. 24, no. 3, pp. 265-272. http://dx.doi.org/10.5511/plantbiotechnology.24.265.

SEMIARTI, E., MERCURIANI, I.S., RIZAL, R., SLAMET, A., UTAMI, B.S., BESTARI, I.A., AZIZ-PURWANTORO., MOELJOPAWIRO, S., JANG, S., MACHIDA, Y. and MACHIDA, C., 2015. Overexpression of PaFT gene in the wild orchid Phalaenopsis amabilis (L.) Blume. *AIP Conference Proceedings*, vol. 1677, no. 1, p. 090005. http://dx.doi.org/10.1063/1.4930750.

STERNBERG, S.H., REDDING, S., JINEK, M., GREENE, E.C. and DOUDNA, J.A., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, vol. 507, no. 7490, pp. 62-67. http://dx.doi.org/10.1038/nature13011. PMid:24476820.

STUDER, G., REMPFER, C., WATERHOUSE, A.M., GUMIENNY, R., HAAS, J. and SCHWEDE, T., 2020. QMEANDisCo: distance constraints applied on model quality estimation. *Bioinformatics*, vol. 36, no. 6, pp. 1765-1771. http://dx.doi.org/10.1093/bioinformatics/btz828. PMid:31697312.

SUN, L. and KAO, T.H., 2018. CRISPR/Cas9-mediated knockout of PiSSK1 reveals the essential role of S-locus F-box protein-containing SCF complexes in recognition of non-self S-RNases during cross-compatible pollination in self-incompatible Petunia inflata. *Plant Reproduction*, vol. 31, no. 2, pp. 129-143. http://dx.doi.org/10.1007/s00497-017-0314-1. PMid:29192328.

TONG, C.-G., WU, F.-H., YUAN, Y.-H., CHEN, Y.-R. and LIN, C.-S., 2020. High-efficiency CRISPR/Cas-based editing of Phalaenopsis orchid MADS genes. *Plant Biotechnology Journal*, vol. 18, no. 4, pp. 889-891. http://dx.doi.org/10.1111/pbi.13264. PMid:31553827.

TSAI, S.Q., ZHENG, Z., NGUYEN, N.T., LIEBERS, M., TOPKAR, V.V., THAPAR, V., WYVEKENS, N., KHAYTER, C., IAFRATE, A.J., LE, L.P., ARYEE, M.J. and JOUNG, J.K., 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology*, vol. 33, no. 2, pp. 187-197. http://dx.doi.org/10.1038/nbt.3117. PMid:25513782.

TURCK, F., FORNARA, F. and COUPLAND, G., 2008. Regulation and identity of florigen: flowering locus T moves centre stage. *Annual Review of Plant Biology*, vol. 59, no. 1, pp. 573-594. http://dx.doi.org/10.1146/annurev.arplant.59.032607.092755. PMid:18444908.

VAIDYA, P., MCDURMON, A., MATTOON, E., KEEFE, M., CARLEY, L., LEE, C.-R., BINGHAM, R. and ANDERSON, J.T., 2018. Ecological causes and consequences of flower color polymorphism in a self-pollinating plant (Boechera stricta). *The New Phytologist*, vol. 218, no. 1, pp. 380-392. http://dx.doi.org/10.1111/nph.14998. PMid:29369384.

WANG, H., RUSSA, M. and QI, L.S., 2016. CRISPR/Cas9 in genome editing and beyond. *Annual Review of Biochemistry*, vol. 85, no. 1, pp. 227-264. http://dx.doi.org/10.1146/annurev-biochem-060815-014607. PMid:27145843.

WANG, S.L., VISWANATH, K.K., TONG, C.G., AN, H.R., JANG, S. and CHEN, F.C., 2019. Floral induction and flower development of orchids. *Frontiers in Plant Science*, vol. 10, p. 1258. http://dx.doi.org/10.3389/fpls.2019.01258. PMid:31649713.

XU, J., KANG, B., NAING, A., BAE, S., KIM, J., KIM, H. and KIM, C., 2020. CRISPR/Cas9-mediated editing of 1-aminocyclopropane-1-carboxylate oxidase1 enhances Petunia flower longevity. *Plant Biotechnology Journal*, vol. 18, no. 1, pp. 287-297. http://dx.doi.org/10.1111/pbi.13197. PMid:31222853.

YU, C.-S., CHEN, Y-C., LU, C.-H. and HWANG, J.-K., 2006. Prediction of protein subcellular localization. *Proteins*, vol. 64, no. 3, pp. 643-651. http://dx.doi.org/10.1002/prot.21018. PMid:16752418.

ZIV, D., ZVIRAN, T., ZEZAK, O., SAMACH, A. and IRIHIMOVITCH, V., 2014. Expression profiling of FLOWERING LOCUS T-like gene in alternate bearing 'hass' avocado trees suggests a role for PaFT in avocado flower induction. *PLoS One*, vol. 9, no. 10, p. e110613. http://dx.doi.org/10.1371/journal.pone.0110613. PMid:25330324.